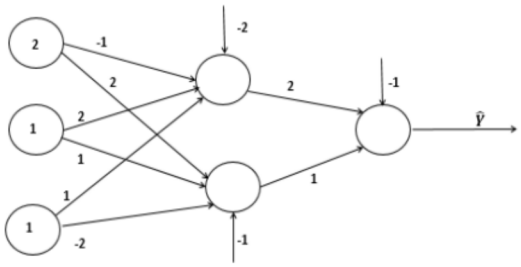
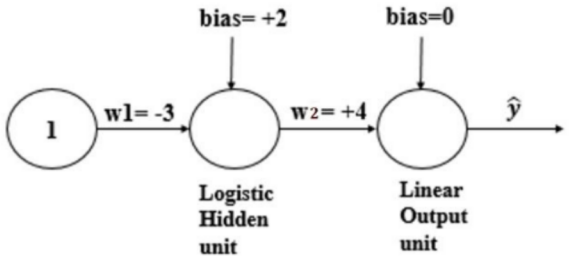
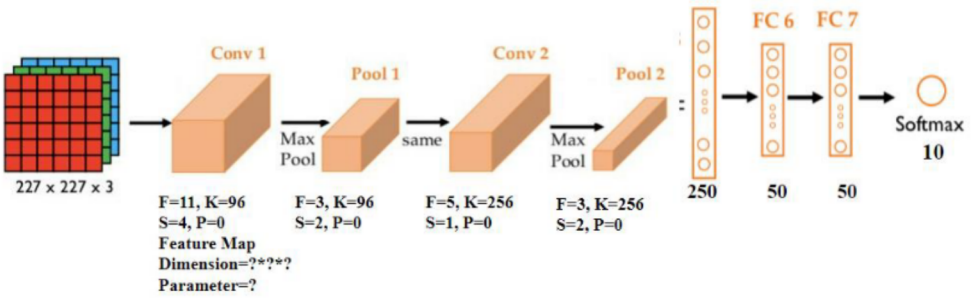
SVKM's
D. J. Sanghvi College of Engineering

Program: B.Tech in Comp. Sci. and Eng.(Data Science) **Academic Year: 2022** **Duration: 3 hours**
Date: 10.01.2023
Time: 10:30 am to 01:30 pm
Subject: Machine Learning- II (Deep Learning) (Semester V) **Marks: 75**

Instructions: Candidates should read carefully the instructions printed on the question paper and on the cover page of the Answer Book, which is provided for their use.
(1) This question paper contains 05 pages.
(2) All Questions are Compulsory.
(3) All questions carry equal marks.
(4) Answer to each new question is to be started on a fresh page.
(5) Figures in the brackets on the right indicate full marks.
(6) Assume suitable data wherever required, but justify it.
(7) Draw the neat labelled diagrams, wherever necessary.

Question No.		Max. Marks
Q1 (a)	1. Answer the following i. Compute the output of the neuron shown below. Transfer function of hidden Layer: Unipolar Hard Limit. Transfer function of output Layer: Binary Sigmoidal  ii. Illustrate mathematically the consequences of not using activation function in every neuron in neural network **OR** 2. Why is the Rectified Linear Unit (ReLU) often preferred over the sigmoid function as an activation function in deep networks? Support your theory with neat labelled diagram and necessary equations.	[05] [05]
Q1 (b)	i. Consider Mc culloch Pitts Neuron for which the inputs are x_1,x_2 and x_3. Also the aggregate function g(x) is an OR function. What is the thresholding parameter for the same? ii. Explain what effect the following operations will have on the (i) bias and (ii) variance of your model. a) Regularizing the weights in a linear/logistic regression model b) Incresing the number of hidden units in an artificial neural model c) Using dropout to train a deep neural network.	[05] [05]

Q2 (a)	1. Answer the following i. Implement perceptron training rule for a network with the following data (X1= [2 1 –1], d1=-1); (X2= [0 –1 -1], d2=1), Initial weights W1 = [0 1 0], repeat the training sequence (X1, d1), (X2, d2) until two correct responses in a row are achieved. Assume c=1 and f(net)=sign(net) ii. What is the condition for convergence of a perceptron learning algorithm? Assume you have a perceptron to solve problem of deciding if a student is eligible for scholarship or not. We have only one input in this case. Bias being 50%. What will be the decision on the model when the student scored 0.49 and 0.51? **OR** 2. Illustrate optimizer functions in Deep Neural Network Model	[07] [03] [10]
Q2 (b)	Consider the neural network given below with one input unit, one hidden unit with logistic activation function, and one output unit with linear activation function. Consider training case of input output pair with input value is 1 and the output value is 1. Suppose Loss function used is $L = \frac{(y - \hat{y})2}{2}$.  a) What is the output of the hidden unit and the output of given training case? b) What is the loss for given training case? c) What is the derivative of loss w.r.t. w1 for given training case? d) Assume you are developing a model to predict the probability as an output. Mention what will be appropriate activation function.	[05]
Q3 (a)	i. Consider the following: Weight array, W= [0.2, 0.7, 0.05, 0.75, 0.86, 0.21] and Input X = [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]. What is the next revised estimate S_t that is obtained by sliding the filter W_t over the input X_t? ii. You are training a neural network model using Early stopping technique. Given that the patience parameter is 2, When will you stop training? Epochs | Training loss | Validation loss 1 | 3.4 | 2.0 2 | 2.0 | 1.9 3 | 1.9 | 1.8 4 | 1.8 | 1.8 5 | 1.7 | 1.9 6 | 1.6 | 2.1	[05]
Q3 (b)	1. Consider the leading CNN architecture for **object-detection** – AlexNet (reduced version) given below. Find out number of training parameters and generated feature map dimension at each layer of CNN. Explain the need and computations at each layer.	[10]



OR

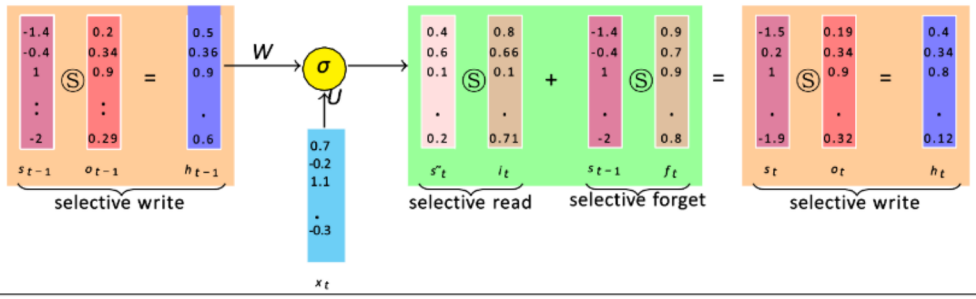
2. How overfitting is handled in deep neural architecture? [10]

Q4 (a)

1. Answer the following
   i. How the RNN, GRU, and LSTM architectures for sequence modelling differ and are comparable. [04]
   ii. Consider the task of Video QA where you want to generate a textual answer. For example, "Q: What is the person in video doing? A: The person in the video is chopping vegetables." Assume all videos are of the same length T and all answers are of the same length J. Here Data = $\{\text{video}, \text{answer}_i\}_{i=1}^{n}$ where video_i =f_{i1},f_{i2},….f_{iT} and $\text{answer}_i = w_{i1}, w_{i2}, \ldots, w_{iT}$, Support your answer with neat labelled encoder and decoder diagram, Write down the equation of the encoder and the decoder. [04]

OR

2. Answer the following
   i. While using prediction-based models to learn word representations, the softmax function is computationally expensive. What are the proposed alternate solutions for the same? [04]
   ii. Given below is the representation of LSTM, What is the number of operations that take place at a given timestep, t? Write the equation for those gates and states of LSTM network. [04]



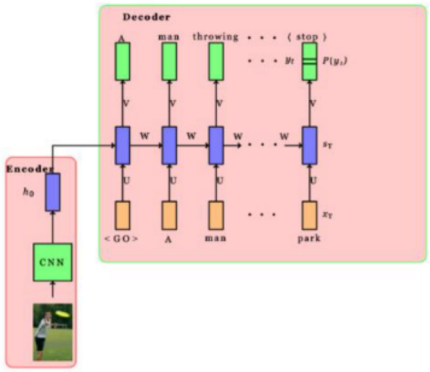
Q4 (b) Imagine C1, C2, and C3 are three chefs and each contribute to the creation of a specific dish. It takes 50 individual steps to complete the dish and they only reach these 50 steps working together. Their individual and combined contributions are divided as follows showing the number of steps each chef completes individually and in combination. [07]

i. Calculate the marginal contribution of each chef depending on the order in which they add their contributions
ii. Calculate Shapley value associated with each chef.
iii. Specify your observation by finding chef who contributed the most towards completing the final dish with a Shapley value.

Chef	Steps Completed
C1	15
C2	20
C3	5
C1 U C2	35
C1 U C3	25
C2 U C3	30
C1 U C2 U C3	50

Q5 (a) **Solve any two.**

1. Consider the following encoder-decoder model of image captioning where an image is given as an input to the model and need to generate captioning of the image. Identify function of the encoder. Suggest better approach of feeding abstract representation of image to the decoder model in order to have better prediction accuracy. Write equation of decoder objective function and activation function. [05]



2. Consider the following corpus: "The event structure constitutes of panel discussion on the theme Data Science the way for future.". [05]
   a) What is the size of the vocabulary of the above corpus?
   b) What is the maximum size of the co-occurrence matrix?
   c) Find Euclidean distance and cosine similarity of the given one-hot representation of the following?

Event:	1	0	0	0	0	0	0	0
Data:	0	0	0	1	0	0	0	0
Science:	0	1	0	0	0	0	0	0

3. In a Generative Adversarial Network, there are two players, the Generator (G) and Discriminator(D). [05]
   a) Sketch basic GAN using the elements correcting to the two players, a source of real images, and a source of randomness.
   b) Write expression corresponding to the optimization objective of Generator G and Discriminator (D).
   c) Mention any one application of GAN in detail.

4. Illustrate working of LIME –Explainable AI Algorithm. [05]

Q5 (b)	Construct a Kohonean Self Organizing Map to cluster the four given vectors [0 0 1 1], [1 0 0 0], [0 1 1 0] and [0 0 0 1]. The number of clusters to be formed is two. Assume an initial rate of 0.5. Initial weight values are randomly initialized as follows, Find weight values after one epoch. $W_{ij} = \begin{bmatrix} 0.2 & 0.9 \\ 0.4 & 0.7 \\ 0.6 & 0.5 \\ 0.8 & 0.3 \end{bmatrix}$	[05]